An integrated, metric physical map and selected sequence from human chromosome 19.

A.V. Carrano, L.K. Ashworth, B. Brandriff, E. Branscomb, E. Garcia, L. Gordon, J. Lamerdin, G. Lennon, H. Mohrenweiser, and A.S. Olsen
Human Genome Center, L-452, Lawrence Livermore National Laboratory, Livermore, California 94551, USA

We have constructed a metric physical map of human chromosome. The foundation of the map are sets of overlapping cosmids (contigs) generated by automated fingerprinting, spanning at least 95% of the euchromatin, about 50 Mb. From these contigs, distances between selected cosmid clones were estimated using fluorescence *in situ* hybridization (FISH) in sperm pronuclei, thereby providing both order and distance between contigs. An average inter-marker separation of 230 Kb has been obtained across the entire non-centromeric portion of the chromosome. Larger insert clones, primarily yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs) and P1-based bacterial artificial chromosomes (PACs), were used to span gaps between contigs. Currently, the map consists of 51 'islands' containing multiple clone types, whose size, order and relative distance are known. Over 450 genes, genetic markers, sequence tagged sites (STSs), anonymous cDNAs, and other markers have been localized to clones contained in the map. In addition, complete digest *Eco*RI restriction maps have been generated for >41 Mb (~83%) of the chromosome. In selected regions of special interest, for example the 29-member CEA gene family, the CYP2A-2B-2F gene cluster, and the fucosyltransferase gene family regions, fine scale restriction maps defining the precise location and orientation of multiple gene loci have been developed.

We have sequenced several hundred Kb of genomic DNA from this chromosome as well as syntenic regions in the mouse targeted primarily to the human DNA repair genes XRCC1 and ERCC2 on chromosome 19, ERCC4 on chromosome 16, and XRCC3 on chromosome 14. We have sequenced 76 Kb containing the human and mouse XRCC1 genes. In addition to the coding regions, 9 conserved elements were identified with sequence identities ranging from 65% to 78%. We have completed 52 Kb of human sequence encompassing the ERCC2 gene as well as 54 Kb spanning the syntenic regions in the mouse and hamster. A defect in ERCC2 leads to the cancer-prone human disorder xeroderma pigmentosum (XP-D). The human ERCC2 gene is comprised of 23 exons and is 98% identical to the rodent homologs at the protein level. We identified two genes flanking ERCC2, one may be a new member of the kinesin gene family, and the other has no known function. Like ERCC2, the ERCC4 gene product is involved in the nucleotide excision repair pathway, which recognizes and removes DNA damage. A total of 35 Kb has been completed for this gene region which has been instrumental in identifying and assembling the coding regions from numerous partial length cDNAs. The full-length gene spans ~29 Kb and is >50% AT-rich. The ERCC4 gene product exhibits significant homology to the S. cerevisiae rad1 and S. pombe rad16 genes, which encode single strand endonucleases. Finally, we have completed sequencing a 2.7 Kb candidate cDNA for the recently cloned human XRCC3 gene and are in the process of sequencing the cosmid containing this gene, which appears to play a crucial role in chromosomal stability.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract no. W-7405-ENG-48.